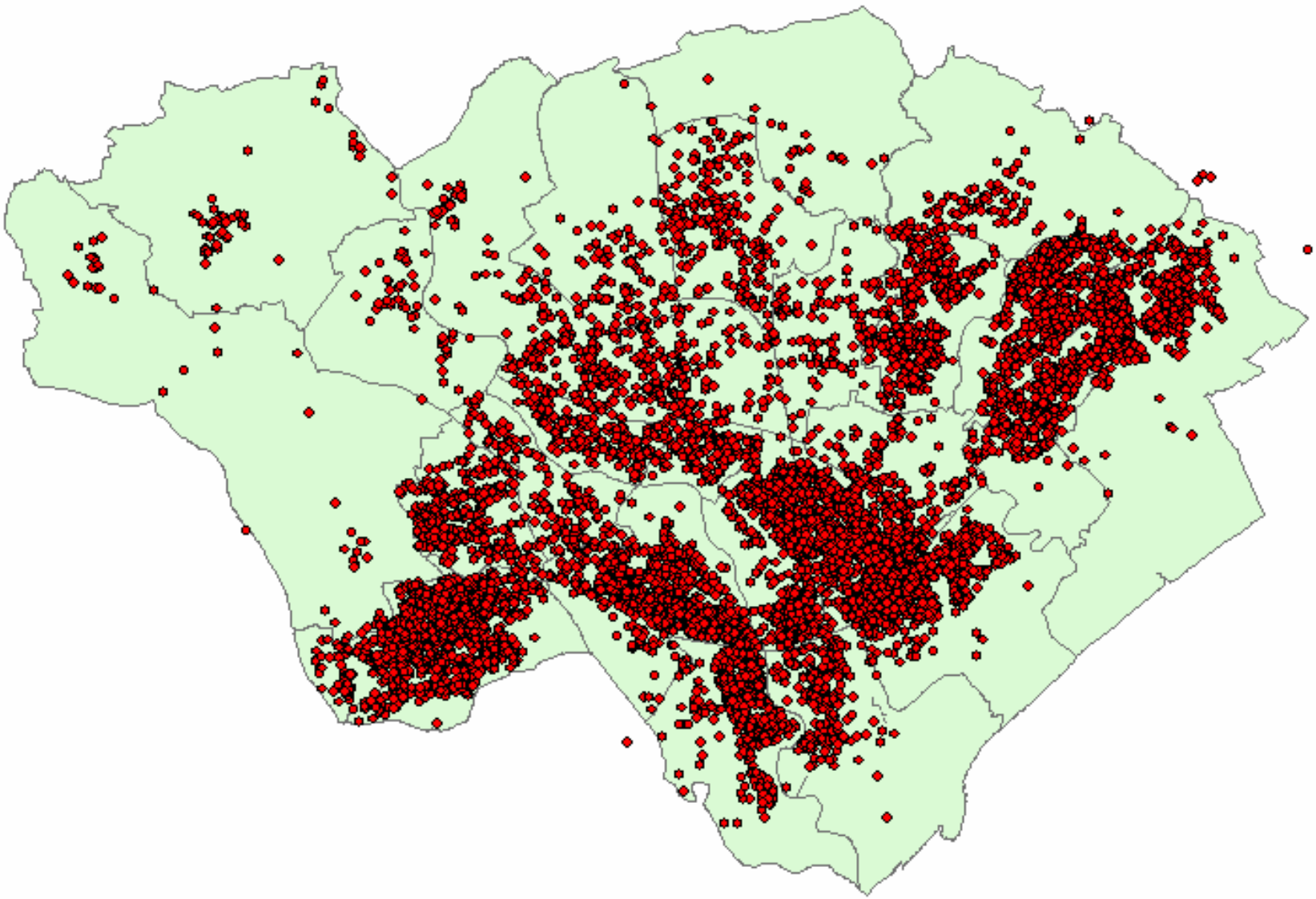
# The Dynamic Spatial Disaggregation Approach

## To Geo-Temporal Crime Forecasting

**Hasan Al-Madfai**, Christian Ivaha and Andrew Ware

# Incidences of criminal damage 2001 - 2004



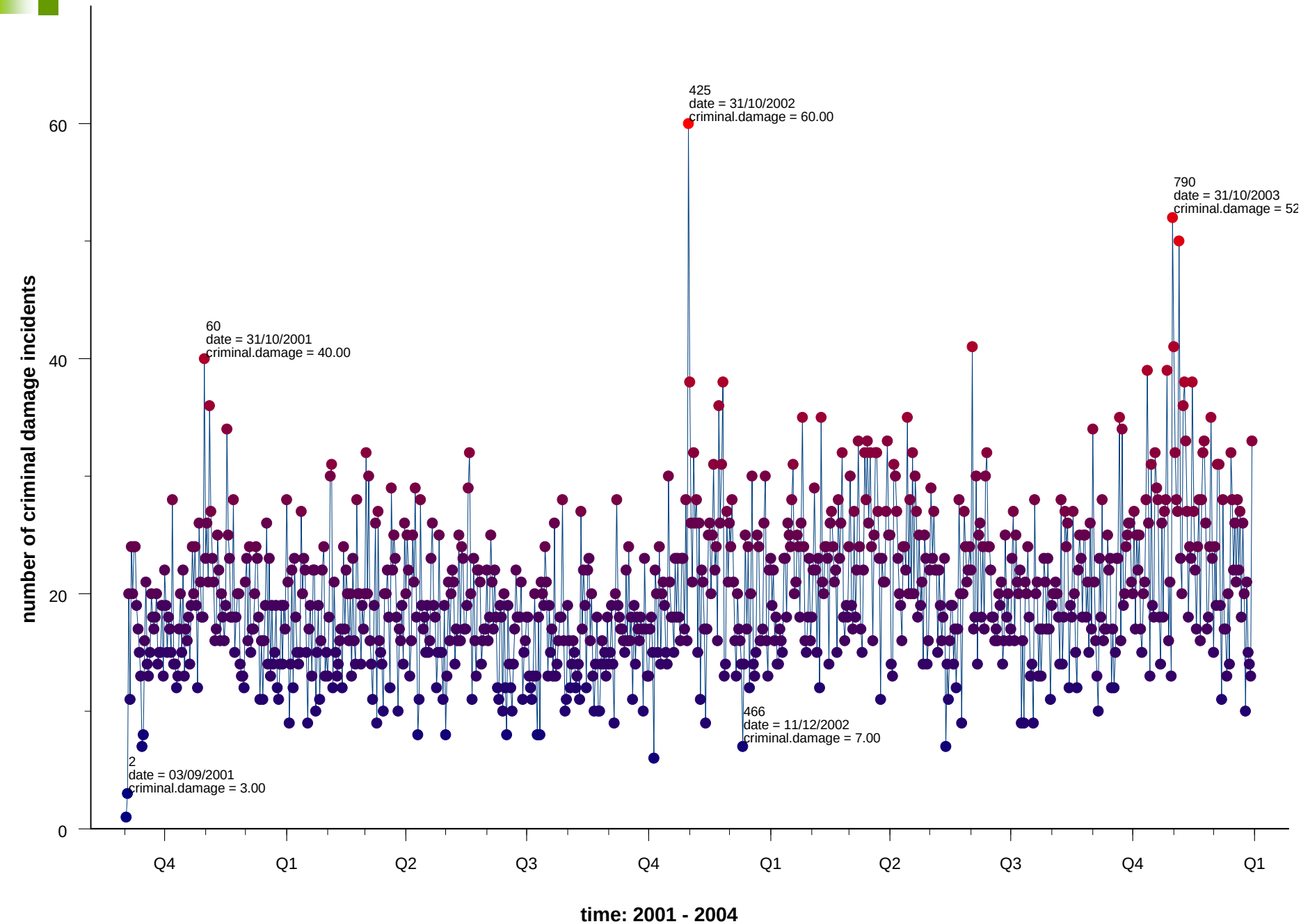
# Motivation

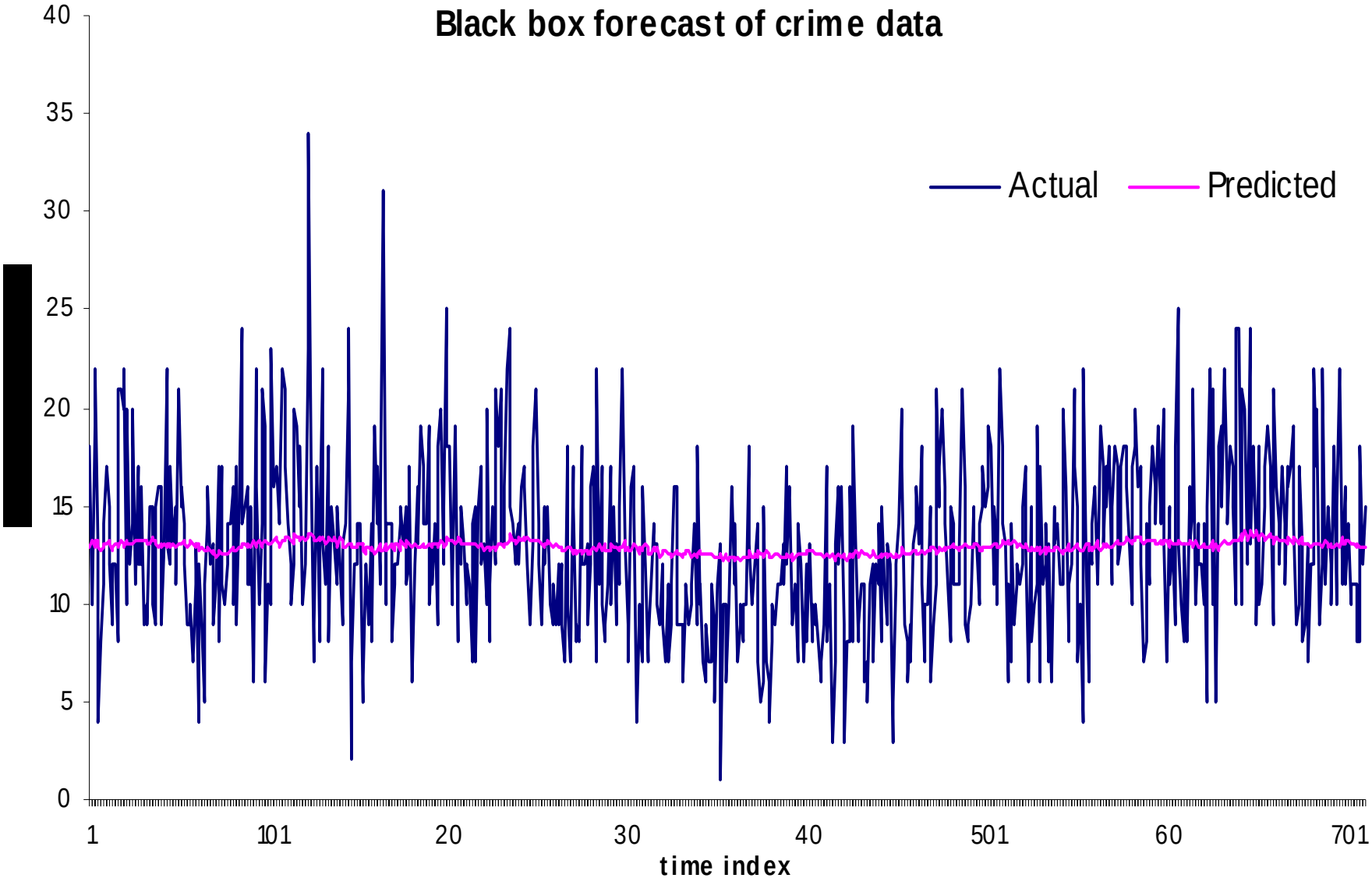
- “How many” incidences are expected to take place
- “Where” are these incidences expected to take place
- “Why” is there variation in the temporal and spatial patterns of incidences of criminal damage

# Paper introduces two innovations:

- The Hierarchical Profiling Approach (HPA) for temporal forecasts
- The Dynamic Spatial Disaggregation Approach (DSDA) for spatial forecasts

# Daily incidences of criminal damage offences




Black box forecast of crime data
Actual
Predicted
40
35
30
25
20
15
10
5
0
1
101
20
30
40
501
60
701
time index

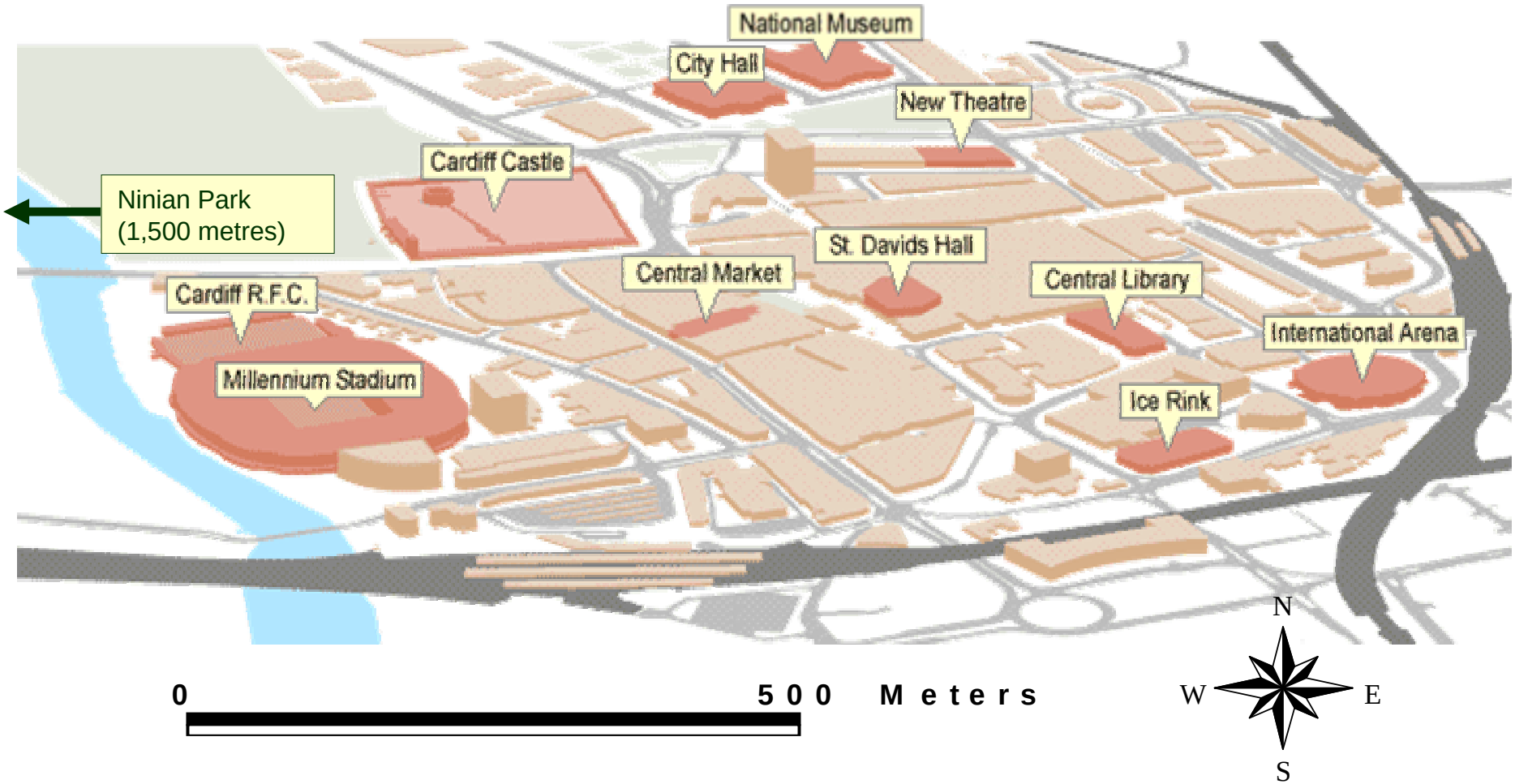
# The Hierarchical Profiling Approach (HPA)

- Uses Profiles to model and correct exotic values, outlying time windows and other periodic and aperiodic disturbances in the data
- Assumes the model

$$y_t = f(t) + Z_t$$

- Where
  - $f(t)$ is a deterministic component that models the profiles
  - $Z_t$ is the stochastic remainder to be modelled.

# Cardiff City Centre




Source: www.cardiff.gov.uk

# Profile estimation

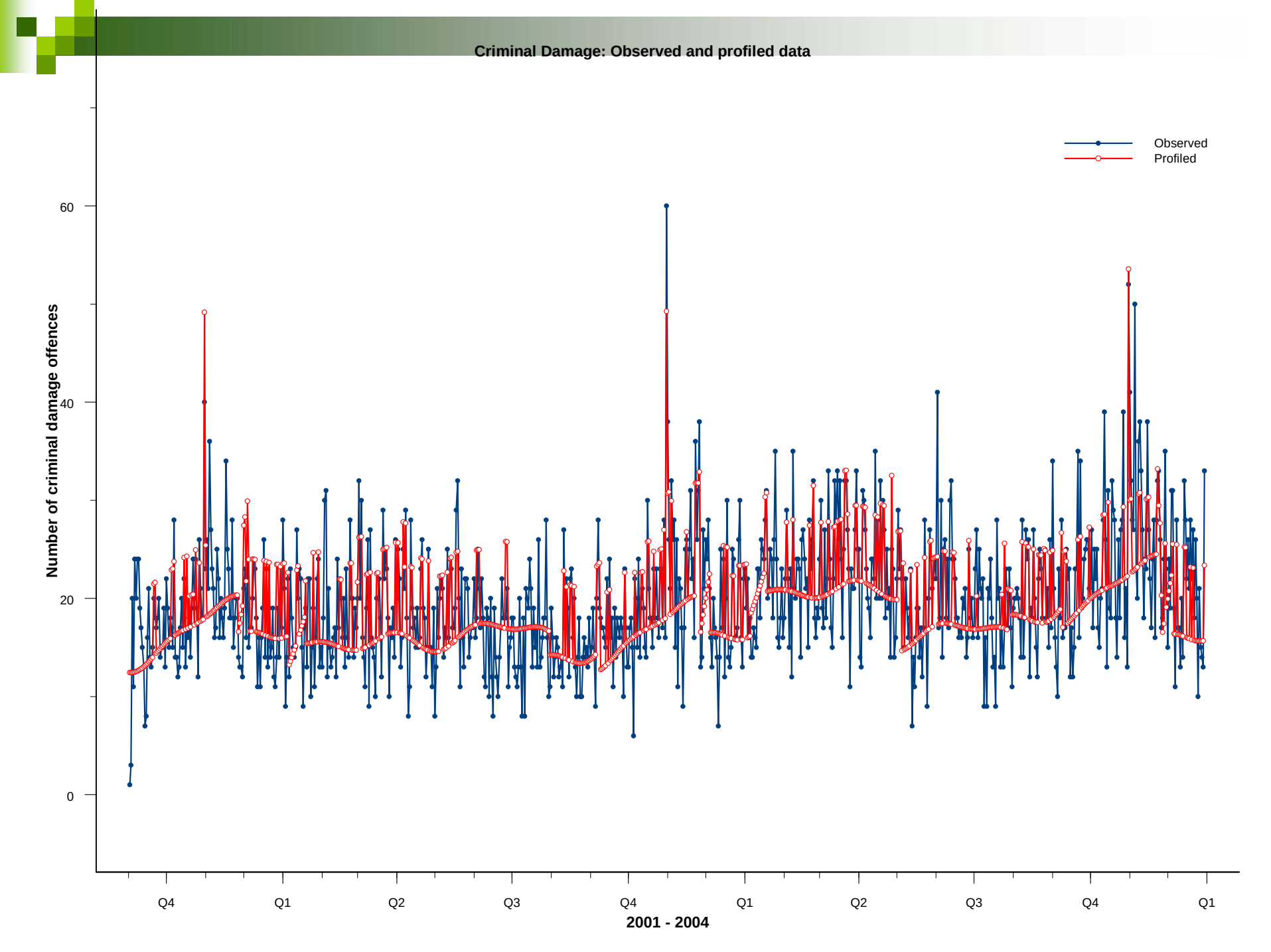
- Profiles can be estimated using aggregated data - metric used is typically OLS
- Need to use continuous functions
- We used
  - OLS point estimates
  - Harmonic Regression:

$$y_t = \sum_{j=0}^{k} \delta_j t^j + \sum_{i=1}^{s/2} \left( \beta_i \cos(t\alpha) + \gamma_i \sin(t\alpha) \right); \quad \alpha = \frac{2\pi}{s}$$

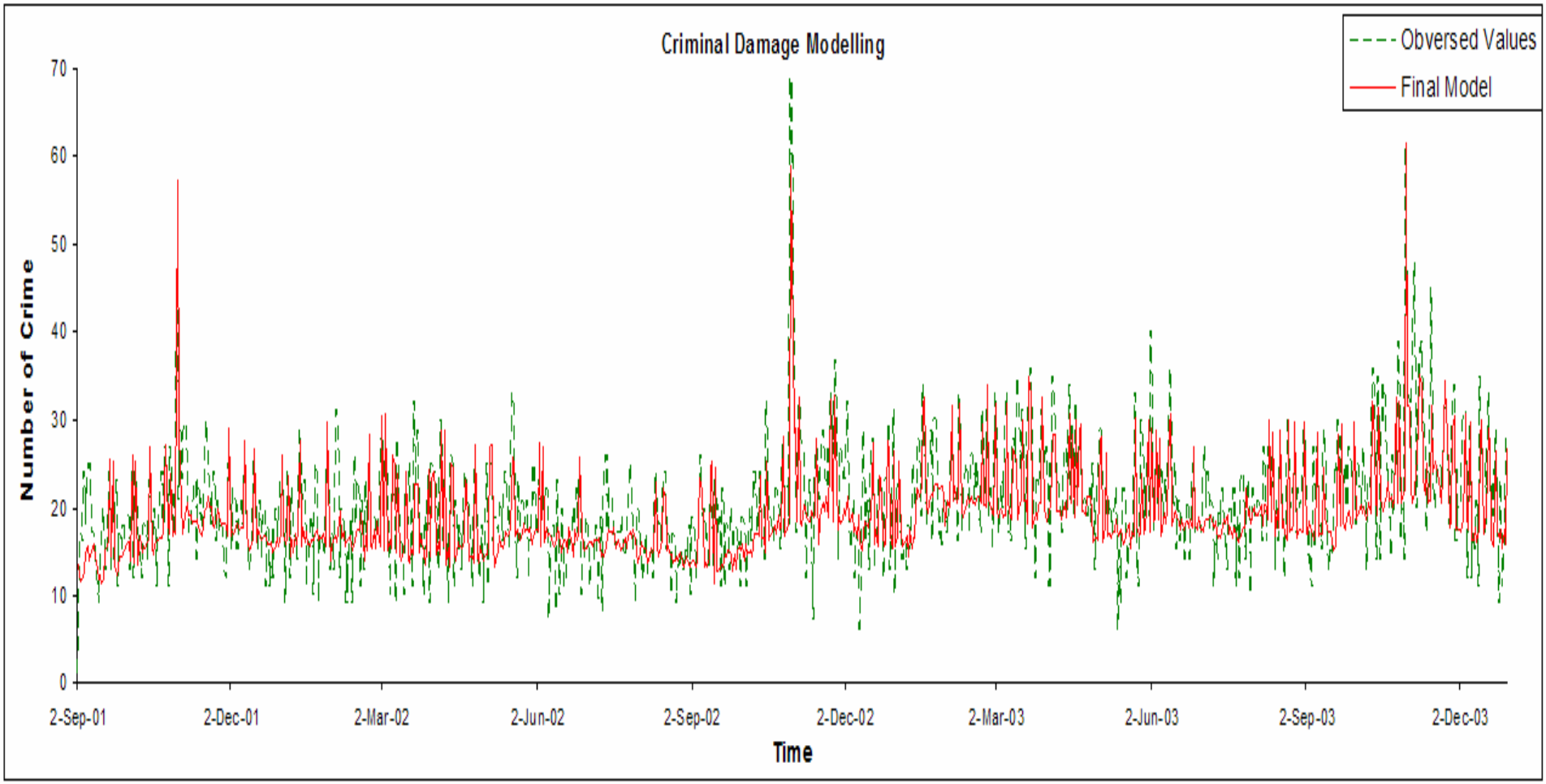
# Selected Level 1HPA Profiles for incidences of criminal damage

| Event Type | Average Deviation |
|---|---|
| Halloween | 32.2 |
| Ninian Park + Rugby Millennium Stadium | 12.6 |
| Six Nation Rugby Games + Rugby Games against New Zealand | 11.4 |
| Easter Weekend | 9.2 |
| CIA Concerts | 9.1 |
| Boxing Match | 9.0 |
| GCCC in May 2002 and Sept 2002 | 8.9 |
| Ninian Park + Football Millennium Stadium | 8.8 |
| Cardiff Union | 8.7 |
| Ninian Park+ Boxing Match | 7.7 |
| Arms Park + Ninian Park + Football Millennium Stadium | 7.7 |

| Event Type | Average Deviation |
|---|---|
| Football Millennium Stadium | 7.6 |
| 1/2 Term / Bank Holiday | 7.6 |
| SpeedWay at Millennium Stadium | 7.4 |
| Ninian Park | 7.3 |
| Rugby Millennium Stadium | 6.9 |
| St Patrick's Day | 6.8 |
| St Patrick's Day + Ninian Park | 6.7 |
| Rugby Arms Park | 6.4 |
| Ninian Park + CIA Concert | 6.2 |
| Ninian Park + Arms Park | 6.0 |
| Cardiff Summer Festival 2003 | 3.4 |
| Other Millennium Stadium Events | 3.2 |

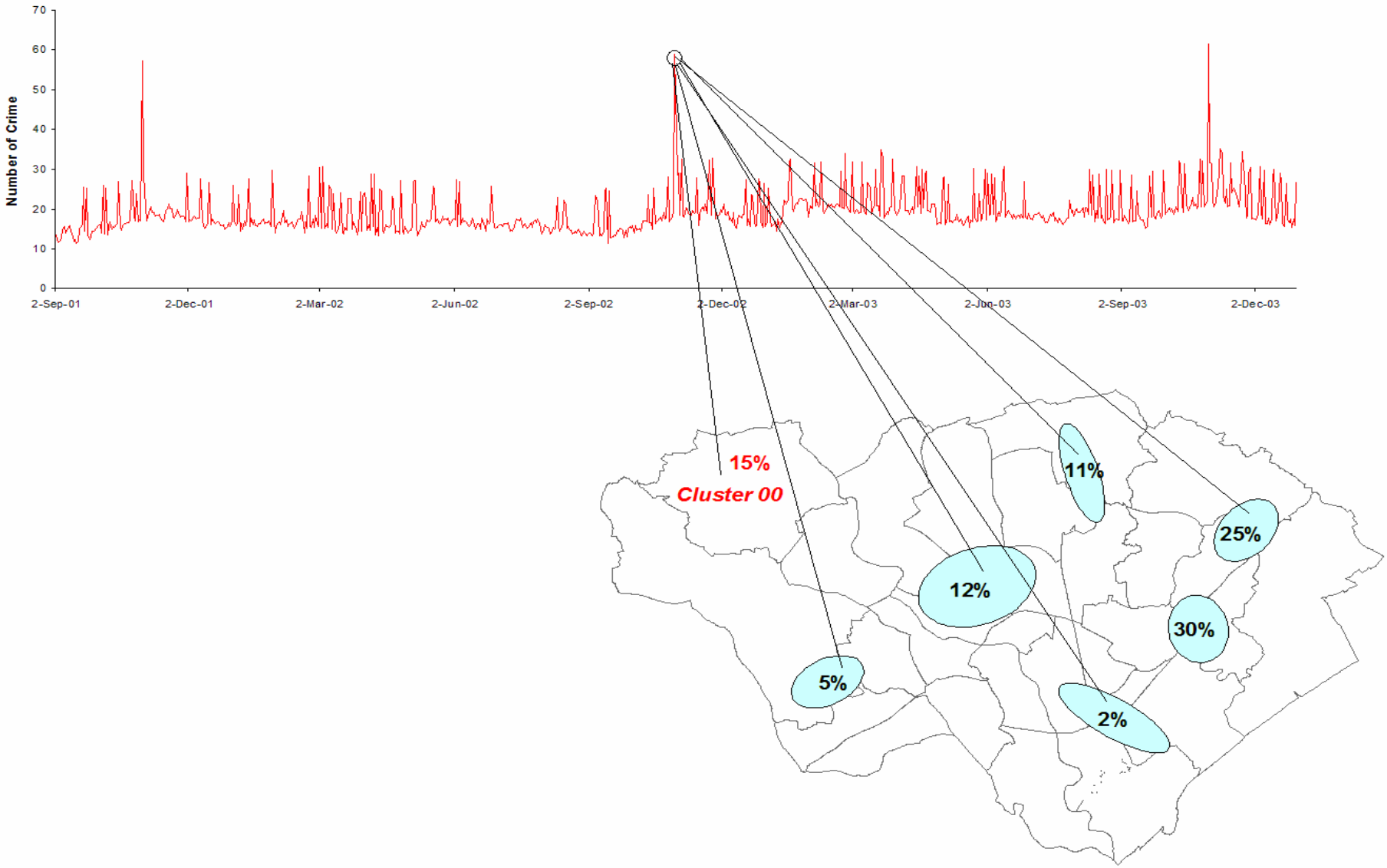

Criminal Damage: Observed and profiled data
Observed
Profiled
Number of criminal damage offences
0
20
40
60
Q4
Q1
Q2
Q3
Q4
Q1
Q2
Q3
Q4
Q1
2001 - 2004

# One Step-Ahead forecasts




Average abs. error: ±3.51 crimes/day

# Dynamic Spatial Disaggregation



# Spatial Disaggregation of Forecasts

- Uses a set of weights, $\beta_{tj}$ called the SFD weights.
- For a target day, *t*, given a forecast $y_{t-1}(1)$, disaggregating the forecasts over *m* clusters is obtained using:

$$\hat{F}_{tj} = \beta_{tj}\left(\frac{y_{t-1}(1)}{m_t}\right); \quad \sum_{\forall j} \beta_{tj} = 1$$

- Main challenges:
  - Forecast Evaluation
  - Determining the SFD weights

# Forecast Evaluation

- The Spatio-Temporal Mean Root Square Error (STMRSE) includes an extra dimension for the spatial element.
- Measured on crime/cluster/day
- Calculated using:

$$STMRSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{m_t}\frac{\left(F_{tj} - \hat{F}_{tj}\right)^2}{m_t}}$$

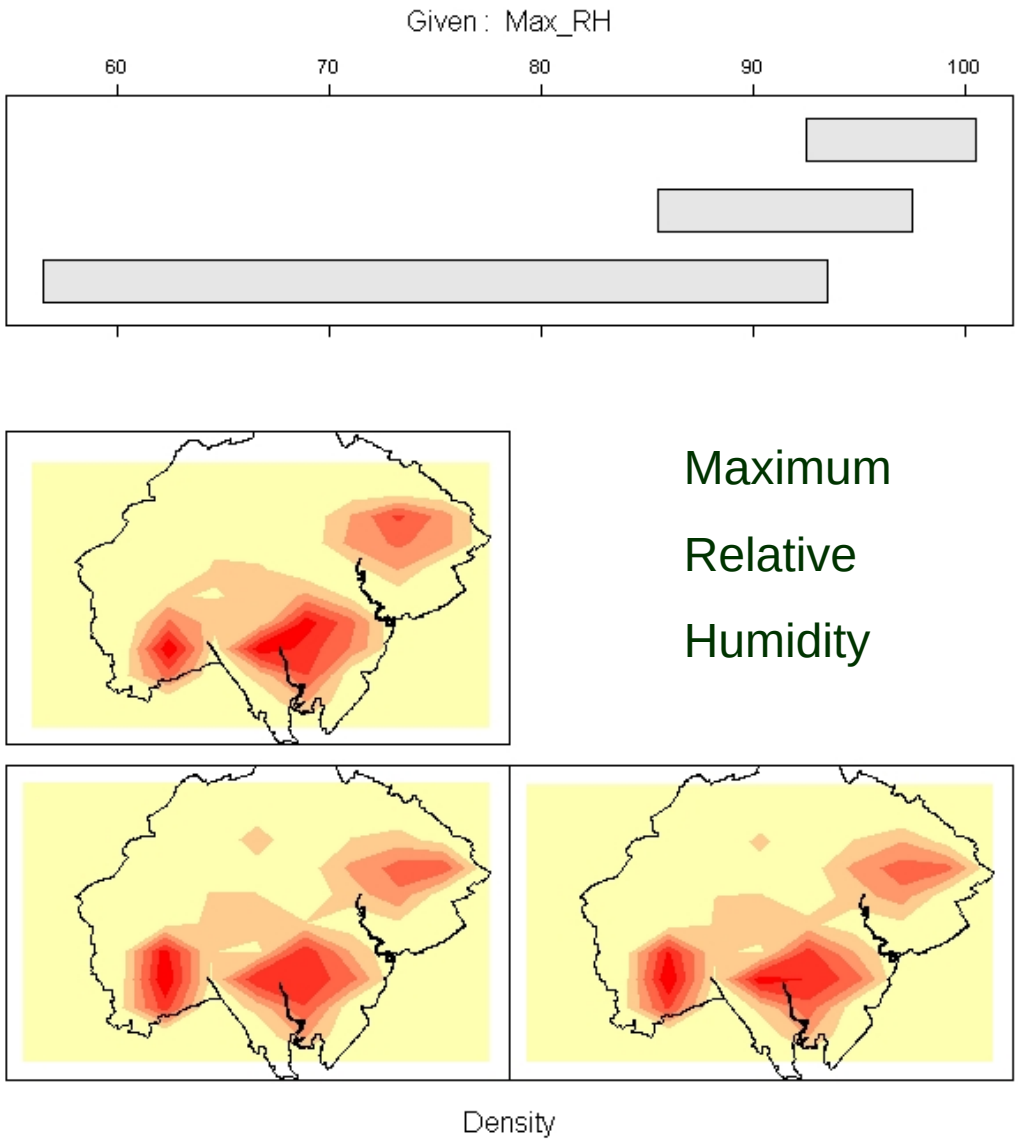
# Determining the SFD weights

- Using Nearest Neighbour Analysis (NNA), there was sufficient evidence to suggest that clusters existed in the dataset ($Z$=-239.2, $p$< 0.01)

- Weekday pattern analysis revealed spatial correlation between weekdays and Census Wards ($\chi^2$=326.12, *d.f =168,  p < 0.05*)

# Naive approach: weekday spatial seasonality

# The influence of ambient conditions




Number of Crime
70
60
50
40
30
20
10
0
2-Sep-01
2-Dec-01
2-Mar-02
2-Jun-02
2-Sep-02
2-Dec-02
2-Mar-03
2-Jun-03
2-Sep-03
2-Dec-03
Time
Cluster 00
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday

# SFD weights beyond the naive model

- Days satisfying each of the identified spatial/ambient conditions isolated – clusters identified using STAC
- SFD weights determined using a number of methods including:
  - OLS. This was used to predict the % of total number of incidence in each cluster
  - Arithmetic mean

# Results

| Naïve approach | Arithmetic Mean | OLS on Percentages |
|---|---|---|
| *1.5892* | *1.1296* | *1.1240* |

STMRSE: Crimes per cluster per day

# Results

| | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday |
|---|---|---|---|---|---|---|---|
| ***Cluster 00*** | 30.16% | 25.84% | 25.51% | 27.90% | 36.03% | 28.71% | 27.00% |
| ***Cluster 01*** | 7.90% | 10.07% | 11.38% | 12.17% | 6.62% | 11.42% | 12.01% |
| ***Cluster 02*** | 9.09% | 8.63% | 10.20% | 7.55% | 5.46% | 8.40% | 7.73% |
| ***Cluster 03*** | 5.99% | 7.80% | 6.05% | 7.04% | 7.74% | 9.61% | 7.60% |
| ***Cluster 04*** | 6.60% | 4.79% | 4.06% | 9.01% | 6.11% | 5.88% | 7.95% |
| ***Cluster 05*** | 5.56% | 7.28% | 9.39% | 6.13% | 3.37% | 4.71% | 4.06% |
| ***Cluster 06*** | 5.91% | 6.10% | 6.37% | 5.08% | 4.10% | 5.92% | 5.01% |
| ***Cluster 07*** | 3.76% | 4.88% | 7.22% | 5.17% | 3.33% | 3.43% | 4.71% |
| ***Cluster 08*** | 3.99% | 3.88% | 3.84% | 3.52% | 3.25% | 4.14% | 4.88% |
| ***Cluster 09*** | 2.19% | 3.05% | 3.70% | 3.89% | 3.25% | 3.58% | 3.93% |
| ***Cluster 10*** | 4.83% | 4.18% | 3.39% | 2.29% | 2.09% | 2.37% | 2.85% |
| ***Cluster 11*** | 2.19% | 2.09% | 2.44% | 1.97% | 2.71% | 2.07% | 4.15% |
| ***Cluster 12*** | 1.84% | 4.84% | 2.66% | 1.83% | 1.97% | 2.15% | 1.94% |
| ***Cluster 13*** | 1.53% | 1.83% | 0.86% | 3.57% | 2.01% | 1.51% | 2.94% |
| ***Cluster 14*** | 3.15% | 2.31% | 1.08% | 2.88% | 1.32% | 3.69% | 1.34% |
| ***Cluster 15*** | 2.65% | 2.44% | 1.85% | | 2.21% | 2.41% | 0.82% |
| ***Cluster 16*** | 1.65% | | | | 1.66% | | 1.08% |
| ***Cluster 17*** | 1.00% | | | | 2.55% | | |
| ***Cluster 18*** | | | | | 1.39% | | |
| ***Cluster 19*** | | | | | 1.51% | | |
| ***Cluster 20*** | | | | | 1.32% | | |

# In conclusion

- The HPA and DSDA are generalised approaches capable of dealing with changing boundaries.
  - Together they give fairly reliable answers to the ‘how many’, ‘where’ and ‘why’ questions of crime forecasting.
- The use of the HPA and DSDA allows for the effects of exogenous variables, on the spatial **and** temporal levels to be quantified and modelled
- One pressing future work item is the identification of the SFD weights
- Improved spatial pattern recognition routines would not go amiss.